Guidelines for Insertion of Commercial Memories into NASA Projects

I. Introduction: Why NASA projects are using increasing amounts of memory

The increasing density and performance of commercial memories relative to their radiation-hardened counterparts have increased the temptation to use these parts until it has begun to border on necessity. Not only does the greater density of commercial memories allow on-orbit storage and processing of large datasets, it also allows the use of commercial software—in which the assumption that memory is cheaper than processing power is more or less axiomatic. While these characteristics can significantly enhance spacecraft capabilities and help projects to meet aggressive schedules, many commercial memory technologies are prone to serious radiation degradation and potentially catastrophic single-event effects. If such hazards in commercial devices are not carefully characterized and mitigated, mission goals may be significantly compromised. Even if such errors are discovered early enough to be corrected, they may result in significant schedule slips.

In this report, we consider how commercial memories can be inserted into spacecraft applications in a way that maximizes their advantages and minimizes their risks. The general approach is illustrated in the flow chart in figure 1. We begin by considering when it may be appropriate to use commercial memories. We then consider the types of commercial memories and devices currently available. While the sheer numbers of technologies under consideration may appear daunting, a few general classifications will prove invaluable in imposing structure on the seeming chaos.

We then consider the radiation effects and other threats to which these technologies may be vulnerable and the mitigation techniques against these effects. Again, although the technologies involved are diverse, the radiation susceptibilities of memories tend to be very similar to those that affect other microelectronics.

Having considered the risks and mitigation techniques involved in using commercial memory, we then consider the process of deciding whether such use is appropriate to the application being considered.

Next, in order to facilitate the selection of promising candidates, we look at criteria that may affect the radiation susceptibility of memories, including device technologies and fabrication and the maturity of the fabrication process. This is perhaps where we see the strongest manifestations of the commercial in commercial memories. In a market where product lifecycles often lend themselves to freshness date codes, rather than long-term qualification efforts, the relative timing of the memory qualification effort and the product life cycle can be crucial.

However, ultimately success in the use of commercial memories depends on the interplay of effective testing to identify radiation vulnerabilities and effective design to mitigate the consequences of those vulnerabilities. In the case of commercial memories, perhaps more than any other part type, design and testing perform a complex pas de deux with application requirements, architecture and mitigation determining the testing that must be done and testing determining the architecture and mitigation necessary to meet requirements. The reason for the added complexity in testing and design is that radiation effects in many memories are so varied and complicated that a test effort unguided by system considerations would likely be both uneconomical and ineffective. Even if the test effort can be limited by judicious design, it will often be expensive and prolonged.

In the final section, we consider strategies for mitigating the high costs of qualifying a commercial memory—as well as for solving other issues like obtaining lot traceability. Such issues are complicated by the fact that even a buy for a constellation of satellites would be considered small by the standards of commercial memory vendors. By banding together—either as a collaboration of individual projects or as a market—users of commercial memory for spacecraft may obtain not just cost sharing and lot traceability, but also, perhaps a degree of cooperation from the memory vendor.

In much of what follows, we will be guided by the International Technology Roadmap for Semiconductors (ITRS)[1]—a remarkable cooperative product of the very competitive semiconductor industry. The ITRS is intended to facilitate and maintain progress in the development of semiconductor devices by identifying common technical, organizational and material obstacles to that progress. Because it is a product of international and commercial

consensus, it does not deal with detailed solutions to these problems. However, as an insight to the development of semiconductor technologies it is invaluable. We also strongly recommend Tim Oldham's 2003 Nuclear and Space Radiation Effects Short Course on scaling and radiation effects. In addition, the other references provide a good sampling of radiation test results for commercial memories. Unfortunately, the references are not exhaustive. Many builders of hardware consider commercial memories to be of strategic importance, and so are unwilling to make their test results public. As a result, I have chosen to draw guidelines from these results rather than quoting them in detail.

II. When to Use Commercial Memory

The main reason for the increased use of commercial memory in space applications is that these parts engender capabilities that cannot be realized with radiation-hardened and space qualified parts. These capabilities may include speed, data retention and low power consumption. However, probably the characteristic of greatest advantage for spaceflight is density: The latest generation of commercial DRAM is 128 times as dense as the densest space qualified counterpart—a 4 Mbit SRAM. The better than hundred-fold savings in board real estate—not to mention the resulting savings in power—may make the radiation and reliability related headaches of using such a part seem worthwhile. Use of commercial memory may confer other advantages. For example, if a satellite is being designed with an eye toward follow-on missions, the rapid pace of development in commercial memory provides a relatively straightforward path to enhanced capability—although radiation performance of a part cannot be predicted on the basis of performance by past generations. Two other seeming advantages—the low cost of commercial memories and their availability with only short lead times—usually prove to be false economies. The cost of qualifying a commercial part and of implementing mitigation against radiation threats usually exceeds the difference in purchase price between those parts and an equivalent rad hard memory. Moreover, the time needed for qualification, design, mitigation implementation and verification can become a threat to project schedule unless the effort is implemented very early in the project and then carefully managed. In many cases, the commercial memory winds up being the strategically important but temperamental diva around which the rest of the system must be built. For this reason, it is important to identify the memory needs early and begin the qualification effort by looking at the types of commercial memories that are available.

III. Types of Commercial Memory

Although novel memory technologies are being introduced at an increasing rate, the devices being introduced can still be classified according to some useful dichotomies. One of these is the distinction between nonvolatile and volatile memory: The contents of nonvolatile memory persist even when the device is unpowered, while volatile memory requires bias (and in some cases much more) to maintain its contents. Regardless of whether the memory elements are volatile or nonvolatile, CMOS support circuitry, which implements the support and logical functions required by the device, usually accounts for a significant minority of the memory chip's area and most of its radiation effects. Because of the very thin profit margins in memory manufacturing, almost all commercial memory (and therefore also the CMOS) is implemented on bulk silicon wafers, making potentially destructive single-event latchup a significant risk for such devices. The support circuitry is also where the most serious logical errors—SEFIs, mode register upsets, etc. occur. Such errors may corrupt the entire contents of the die in which they occur. Thus, regardless of the specific technology of the memory array, many of the most serious radiation effects are common.

A. Nonvolatile Memory

In some nonvolatile memories (e.g. PROMs), the contents are programmed once and stored permanently. In others (EEPROM, FLASH, etc.) the contents are programmed and stored until they are reprogrammed. Currently, the nonvolatile memory market is dominated by FLASH (NOR or NAND), although FeRAM, magnetoresistive technologies, SONOS and phase-change (e.g Chalcogenide) technologies are thought to be on the near horizon. An interesting aspect of nonvolatile memories is that the memory cell often exhibits significant immunity to radiation effects, while the support circuit often forms an Achilles heel for space flight applications. For example in FLASH technologies, the memory cells are generally immune to SEE if not being Read or Written. Unfortunately,

FLASH charge pumps have tended to be quite susceptible to damage at low TID levels, making Flash unsuitable for all but the most benign radiation environments. Similarly, the early FeRAM designs exhibited good immunity to upsets and TID degradation, but the CMOS support circuitry was susceptible to SEL.

However, while nonvolatile memory cells are often hard to SEE, this cannot be taken for granted. The floating gates of FLASH and other technologies may exhibit susceptibilities to radiation induced leakage current. Likewise, since data in SONOS memories are stored as trapped charge in the nitride layer, one would expect radiation events that affect such stored charge to affect data retention. While technology may give clues as to radiation susceptibilities, only appropriate testing can answer such questions definitively.

B. Volatile Memory (RAM)

For volatile memory, the main distinction is between Static Random Access Memory (SRAM), in which the storage elements are bistable (i.e. 0 as stable a 1) and Dynamic Random Access Memory (DRAM), in which the storage elements are not stable, but must be refreshed periodically to maintain their contents. The simplest SRAM can be implemented as two coupled inverters, or 4 transistors (a so-called 4T cell). However, the stability can be improved significantly by adding other elements, with the most common design having six transistors (6T). SEE hardened SRAM cells often employ still more transistors (e.g. 10T designs), with SEE hits being required in two or more transistors being required to upset the cell. It is interesting that while these designs have proved to be quite immune to SEE for technology nodes with feature sizes in the micron and submicron range, at the 90 nm node and below, the size of the ion track is of the same order as the cell size, making upsets in these cells much more frequent. This development demonstrates that hardening techniques will not necessarily scale with the underlying CMOS technology.

The DRAM memory cell is usually implemented as a single capacitor charge storage element with an access FET restricting flow of current onto and off of the storage element. But the simplicity of the DRAM cell comes at a cost. Because it is impossible for a FET to have zero leakage current, the information on the storage element can be maintained only temporarily (~64 ms) and must be read and refreshed on a periodic basis. Generally, it is the complex CMOS support circuitry needed to support the Refresh functionality, as well as the circuitry needed to organize and maintain such a complex data array (mode registers, etc.) that are responsible for some of the most serious radiation effects in DRAMs (SEL, SEFI, block data errors, etc.). Radiation effects in the memory array tend to be more localized—single- and multi-bit SEU and stuck bits—or global—for example, degraded retention time due to TID induced leakage current in the access FETs. Because of the stringent limits on access FET leakage current needed to achieve adequate data storage, often these devices are the first to degrade or fail due to TID.

C. Hybrids and Stacking

One of the most common reasons for using commercial memories is the need for high memory density—that is, flying the most memory in the least space and for the least weight. In many cases, the space available for the memory is too limited even for commercial memories. In such cases, one may opt for a stacked memory—multiple memory die or even packaged devices implemented as an integral packaged part with a footprint roughly that of a single packaged memory. In some commercial applications, stacked parts are even thinned to a few tens of microns before stacking. Indeed, there is no reason why all of the parts in a single package must be memories. Hybrid technologies may be implemented with a variety of functions in addition to memory. It may even be possible to implement a memory and the mitigation of its radiation effects within a single packaged part.

While such parts may significantly simplify the tasks of the designer, they can also pose new and significant risks. Inclusion of multiple parts—especially memories—in the same package may increase thermal stresses on all of the parts, and many of the most serious risks to memories (especially SEL susceptibility) worsen with increased temperature. Moreover, unless radiation effects are mitigated within the hybrid, inclusion of multiple

memories makes it difficult to mitigate radiation-induced and random failures, since isolating the failed part from the others in a package may not be possible. This means that stacked memories will often also fail as stacks—further increasing the loss of data or functionality.

Thinning of die to maximize density poses further problems. Not only is this likely to further exacerbate thermal stresses—both by increasing the numbers of memories and decreasing the thermal masses of these parts—it may also change the charge collection characteristics of the part, especially for error modes such as SEL.

IV. Radiation Effects and their Mitigation

To better understand the mitigation strategies employed for commercial memories in space systems, it is useful to understand the radiation effects to which these are vulnerable and their possible consequences. For the most part, radiation effects in commercial memories are at least qualitatively the same as those in other microelectronics—at least in part because, regardless of the memory cell technology, the control circuitry is CMOS. What is different for commercial memories is the complexity of the effects exhibited. Commercial memories may exhibit multiple single-event latchup (SEL) modes—some probable, some not; some destructive and some nondestructive. The complexity of the control logic all but ensures that commercial memories will exhibit a variety of upset and SEFI modes, and the small size of memory cells has made single-bit upsets less common than multi-bit upsets. Another manifestation of the small sizes of memory cells is the increasing susceptibility of some technologies (notably DRAMs) to stuck bits—localized deposition of charge in device oxides that renders the associated bit unprogrammable. In DRAMs especially, the thin oxides, coupled with the high performance demands on access FETs renders them especially susceptible to the occurrence of stuck bits. At the same time, the thin oxides may facilitate annealing of holes responsible for the stuck bits by tunneling of electrons into the oxide. To date, the result has been that stuck bits have tended to play a minor role for most memories, but in some systems, they may be responsible for the majority of data errors by end of life (EOL). Because of the complexity of the radiation effects seen in commercial memories, most successful mitigation schemes tend to ignore detailed understanding of each and every SEL or SEFI mode, and instead mitigate the consequences. In part because it may not be possible to find a part that is immune to serious radiation effects, and in part because susceptibility to radiation effects may vary considerably from lot to lot (and perhaps from part to part), most successful schemes adopt a multi-tiered defense against radiation effects. Selection of the best performing parts based on test data forms the first line of defense, but subsequent lines of defense play a crucial role not just in improving memory performance, but also in increasing the probability *a priori* that testing will yield at least some parts that will perform adequately. In what follows, we consider the defenses used against degradation mechanism, destructive SEE and data loss due to SEUs, block errors and SEFIs.

1) Degradation Mechanisms—To date, commercial memories have been majority-carrier technologies, so the predominant degradation is total-ionizing dose (TID). TID degradation can be viewed as a failure mechanism as a result of exposure to a stress (the TID). One way of decreasing the probability is to decrease the stress by adding shielding. The main question is how much shielding one must add to ensure adequate probability of success. Extrapolating performance of a large flight lot from a small test sample is a fraught proposition, since a small flight lot is unlikely to detect pathologies such as bimodal radiation response. In some cases more than an order of magnitude lot-to-lot variation has been seen. If the memories being used have no lot traceability, or if pathological radiation behavior is suspected, it may pay to take a more conservative approach, shielding the parts to the lowest dose possible (well below the test part failure level). One other strategy for increasing reliability in the face of uncertainty over memory performance is increased redundancy. Often the applications most likely to use large numbers of commercial memories (e.g. solid-state recorders (SSRs)) may be able to implement a degree of redundancy. Even though the spare parts will accumulate TID at the same rate as the active parts, the addition of redundant parts means that system failure is no longer driven by the worst-performing part. As redundancy approaches 2:1,

the system performance approaches that of the median of the failure distribution, and even a small sample size has a reasonable chance of predicting system performance. Because redundant systems are inherently more complicated than nonredundant ones, care must be taken in implementing redundancy to ensure that net reliability is increased.

2) Destructive Failures—As mentioned above, SEL is the predominant destructive SEE in commercial memories. Although one could view SEL as a failure mode due to exposure to a stress (the ion LET), shielding is ineffective against SEE, due to the penetrating nature of galactic cosmic rays (GCR). However, because SEL rates increase with temperature, it may be possible to minimize SEL susceptibility by ensuring that parts are flown in a temperature environment where SEL occurs only at negligible rates. (Note: SEL rates also increase with supply voltage, but it is usually not practical to fly the memories with below-nominal supply voltages.) Again, the answer to the question of how low the temperature must be maintained is not necessarily simple. SEL susceptibilities have also been seen to vary significantly from lot to lot and even part to part for commercial memories (especially SDRAMs). Certainly, it is advisable to maintain the temperature below the temperature at which SEL is first observed or as low as possible. Moreover, since most random failure mechanisms are also thermally activated, low-temperature operation can only improve memory system reliability.

   If SEL rates are too high to be negligible on the timescale of the mission, the only alternative is to mitigate the effects of the SEL. One strategy that has worked in the past is to employ overcurrent detection and protection circuitry. Usually, this involves sensing the current rise due to an SEL (usually as a voltage drop across a resistor) and cycling power to the device before any damage occurs. There are several potential issues with implementing such circuitry. Because the current sensing threshold must be low in order to allow rapid power cycling, such circuits are inherently susceptible to single-event transients in the sensing circuitry manifesting as spurious SEL indications These spurious shutdowns may be a significant nuisance during normal operations. Another problem is the difficulty of implementing such circuitry for stacked parts and hybrids, since such parts do not usually allow access to supply currents for individual die within the stack/hybrid. However, perhaps the most difficult issue for this strategy is that of detecting whether latent damage may occur despite the SEL detection/protection circuitry. This risk is further exacerbated if a part exhibits multiple SEL modes. Often, it is too difficult and costly to eliminate latent damage as a concern, and all SEL modes are assumed to be potentially destructive. In this case, the most effective strategy may be to implement redundancy by flying cold spares. At least potentially, cold spares can significantly improve system reliability, since they are not vulnerable to SEL until they are swapped for a failed part and powered up. However, again, care is essential if the memory system is to be implemented in a manner that maximizes its reliability.

3) Data Loss—In principle, data errors may occur in two main ways, by upsets in the memory cells themselves or by upsets in CMOS support circuitry that disturb normal operations sufficiently to compromise data integrity. Due to the small sizes of current microelectronic devices, a single ion may corrupt anywhere from zero (some memory cells are immune to upset) to dozens of individual memory cells. However, SEFIs in the support circuitry may corrupt every word stored on the particular die in which the SEFI occurs. Thus, the amount of mitigation required will depend on the application requirements as well as the memory technology, and the particular memory being used. In some applications, occasional loss of even large chunks of data due to a SEFI may be acceptable, while in other applications, the consequences of any error are unacceptable, and the error rate must be driven as low as possible. Of course, the preferred method of error reduction is to start with a memory that has acceptable immunity to SEE, but this cannot be guaranteed. The techniques discussed allow system performance to be optimized even in the face of inherent weaknesses in performance of the parts that make it up. One of the most important and flexible of these techniques is the use of error correction code (ECC). ECC uses a series of coded bits for each word stored in memory to identify and correct corrupted bits in that word. The rudimentary example that illustrates how such a code might work is the use of a parity bit, which can detect, though

not correct, any odd number of upsets in a word. A Hamming code can correct a single-bit and detect up to 2 corrupted bits. More powerful codes, such as Reed–Solomon algorithms can correct two and detect up to 3 corrupted bits. With some additional over overhead, these codes can also be implemented at the nibble level (that is, 4 logically adjacent bits). Given the small sizes of memory cells in commercial memories, implementation of ECC at the nibble-level might seem to be indispensable. However, in many memories, logically adjacent bits are not physically adjacent. In such memories, a single event may upset many bits, but they may all be in different words, and so still correctable with ECC. Unfortunately, the logical-to-physical map is unlikely to be known a priori during the planning stages for the memory, and the small feature sizes of state-of-the-art commercial memories can make determination of this map exceptionally difficult even with test data. Thus, even though a nibble-based correction scheme may not be strictly necessary to correct multi-bit upsets for the memory that is selected, it may be advisable in order to keep the choices as broad as possible. Moreover, in conjunction with memory system architecture, a nibble-based ECC makes it possible to implement mitigation of SEFIs.

The sheer magnitude of the numbers of errors likely to be generated by a SEFI or other block-data error is likely to overwhelm any ECC algorithm—at least if the algorithm is employed in isolation. One way to mitigate SEFIs is to implement a triplicate-and-vote scheme, with each word stored on three independent memory devices. While, in principle at least, such a system can be made bullet-proof for single errors, upsets in the voting circuitry form a potential Achilles heel of such a scheme. Moreover, the 3x increase in memory and increased system complexity for such a scheme may be unacceptable in many applications. Another approach is to spread each logical word over several memories. For instance, a system using a nibble-correct Hamming code and storing a single nibble from each word on any single die would be immune to any error that affected only a single die. A double-nibble-correct Reed–Solomon code would require SEFIs or upsets affecting the same word in three independent memories—an exceedingly unlikely occurrence. Again, while this might seem to be a recipe for a "bullet-proof" memory, there are caveats. The first caveat is that the radiation environment can shoot a lot of different kinds of bullets. Stuck bits look like permanent SEUs to the ECC and decrease its effectiveness for the affected word. Potentially more serious, if a destructive SEE, such as SEL, causes a part to fail, every single word in memory is affected. Thus, if such a memory architecture is to be implemented, it is extremely important that the memory used be free of destructive failure modes or that these failures be adequately mitigated.

In addition to the above mitigation techniques, there may be strategies that are effective for a particular technology. For example, in DRAMs the data retention may degrade as a result of TID or of locally deposited dose (an almost-stuck bit). Data retention can be improved by increasing the refresh rate, although this measure may be impractical beyond a certain point.

The above techniques provide a means of obtaining acceptable radiation performance from commercial memories that may themselves be inconsistent or marginal. By accepting a priori the need to mitigate the radiation performance of these parts and making provisions for mitigation, one limits the testing that must be done and increases the odds that testing will yield acceptable parts. One also immunizes the system against the possibility that flight-lot performance could fall somewhat short of the qualification lot. In the next sections, we consider first selection of potential candidates for testing and the tradesoffs between testing and design.

V.

VI. Trends in Commercial Memory Development

The major forces driving development of commercial memory are scaling and economics. DRAM minimum feature sizes are among the metrics enumerated by the National Semiconductor Roadmap—the unofficial guideline for remaining competitive in the semiconductor marketplace. This has often meant that when innovations appear in the semiconductor market, they are often

introduced first in memories—especially SDRAMs—and that memories often have the shortest product cycles—as little as 18 months. The drive toward innovation also means that profit margins are slim, and memory manufacturers are among the first vendors to be affected by a downturn in the semiconductor market. This vulnerability in turn drives two other trends in the memory market—consolidation and a variability of product cycle with business cycle. The implications of these cycles and trends for memory performance, supply and qualification are discussed in the next section.

A. Product Cycles and Business Cycles

As is natural in a commercial, the forces that drive memory development are primarily economic: The need to be competitive drives innovation, while the need to hold down costs in a very competitive marketplace pushes toward slower development and a more stable product line, optimized for yield and electrical performance. These competing pressures result in several trends and characteristics that differentiate commercial from the military/space suppliers.

Many of the advantages are well understood—lower power, higher speed, densities up to 256x those attainable with radiation-hardened parts, and so on. Other advantages may be more subtle or specialized. For instance, new generations of parts may be near drop-in replacements for previous generations because this reduces the costs associated with package qualification by the vendor—a characteristic that may be especially advantageous for follow-on missions.

Many of the disadvantages are also clear. Perhaps the most significant disadvantage of commercial memories is that of obtaining lot traceability—which is not available as part of the standard commercial product flow. Rather, commercial facilities rely on statistical methods to optimize production—primarily in terms of electrical performance and yield. Unfortunately, since radiation performance is not part of this optimization equation, commercial memories may be especially variable from lot to lot for some radiation effects. Moreover, because inserting a lot-traceability requirement would significantly disrupt product flow, most vendors will not accept such a requirement. At best, it may be possible to persuade a vendor to sell an entire wafer before dicing and packaging. This poses significant issues. It requires having sufficient confidence that qualification based on generic (i.e. non-lot specific) test data is representative that one is willing to make the substantial investment in a full wafer. Since response to many radiation threats may vary significantly from lot to lot, it may be difficult to achieve such a level of confidence without a substantial investment in testing—or significant collaboration and data sharing.

The difficulty of qualifying a part for use in space is further exacerbated by the relatively short product lifecycles of commercial memories. At its fastest pace, semiconductor development can take a part (or at least a revision of a part) from introduction to obsolescence in 18 months. However, as indicated in figure 1, cycle time can be significantly slower during downturns in the electronics industry, when there is little incentive to innovate.

Thus, from the point of view of space systems, the tail end of a downturn may be the best time to select memories.

B. Commercial Memory: The Bleeding Edge of Technology

1. **Semiconductor Roadmap**
2. **Scaling Challenges And New Materials, Structures and Processes**

C. Radiation Issues in Commercial Memories

1. **Destructive Single-Event Effects**
2. **Nondestructive Single-Event Effects**
3. **Degradation Mechanisms**
4. **Stuck Bits**

VII. Memory Decision Tree

A. Whether to Use Commercial Memory

B. What Kind of Commercial Memory to Use

C. Developing a Memory Architecture

VIII. Design vs. Qualification Cost-Benefit Tradeoff

A. Radiation Effects and Mitigation Techniques